FINAL PROJECT

STAT 515: Applied Statistics and Visualization for Analytics

Regional Price Parity for each state

Submitted By

GROUP 8

AAKASH BOENAL	**G01413569**
MEGHANA KATTA	**G01408184**
RAHUL KOMARAVELLY	**G01422037**

Data Analytics Engineering

Professor Tokunbo Fadahunsi

May 12, 2023

TABLE OF CONTENTS

ABSTRACT

The goal of this project is to study the Regional Price Parity (RPP) dataset to examine the variations in the cost of living across different states in the US. By studying the relationship between RPP and GDP, we aim to gain insights into economic dynamics and their implications for various stakeholders. This project utilizes data from the Bureau of Economic Analysis (BEA) and employs visualizations, including scatter plots, regression lines, heatmaps, and micromaps, to present key findings and facilitate a comprehensive understanding of regional price parities.

1. INTRODUCTION

The Regional Price Parity (RPP) dataset offers insight into the cost-of-living differences among states and metro areas in the United States. This information is valuable to policymakers, researchers, and individuals who want to understand how the economy affects different aspects of life.

2. METHODOLOGY

2.1. DATA COLLECTION

We sourced the RPP dataset from the Bureau of Economic Analysis (BEA) to ensure accuracy and reliability. The dataset includes price parity scores for each state, allowing us to compare the cost of living against the national average.

2.2. DATA ANALYSIS

We conducted data analysis by utilizing various visualization techniques, including scatter plots, regression lines, and micromaps. These visualizations provide a clear representation of the relationship between RPP, GDP, and other relevant economic indicators.

3. DATASET

States	RPP	GDP	Personal Income
Alabama	88.139	209979.3	246873.9
Alaska	104.439	50869.4	40052.1
Arizona	96.721	347656	362113.5
Arkansas	89.445	123347.3	148568.7
California	111.797	2874730.8	2332657.6
Colorado	103.009	373763.3	346082.1
Connecticut	102.603	246555.9	253919.7
Delaware	97.677	64404.7	53406
District of Columbia	111.271	126983	50398.4
Florida	101.43	1029575.6	1160013.8
Georgia	95.784	575292.2	545636.2
Hawaii	119.227	74547.2	67312.3
Idaho	91.776	80093.8	94097.1
Illinois	101.412	780060.8	728882.6
Indiana	92.735	346240.9	359702.4
Iowa	89.568	179753	176780.1
Kansas	91.157	162290.9	164557.3
Kentucky	89.124	197818.3	225018.8
Louisiana	91.276	221152.7	238268.8
Maine	97.205	63594.5	71621.7
Maryland	106.223	368571.1	351518.8
Massachusetts	106.555	533102.1	475686.6
Michigan	94.253	481778	522600.8
Minnesota	98.423	346204.3	333417.9
Mississippi	86.601	104353.5	135579.6
Missouri	92.022	295687.3	321701.8
Montana	91.567	48976.2	59577.9
Nebraska	91.751	122136.1	113636.3
Nevada	95.543	159567.3	171884.2
New Hampshire	102.51	82986.3	86042.2
New Jersey	109.099	566893.2	567508.6
New Mexico	89.907	93625.1	102712.2
New York	109.504	1514779.2	1207418.8
North Carolina	93.805	541933.8	548116.5
North Dakota	91.103	53803.6	47613.4
Ohio	92.459	629287	628659.7
Oklahoma	90.269	193230	206387.4
Oregon	103.032	227979.1	220212.2
Pennsylvania	96.371	710973.1	750118.9
Rhode Island	102.083	54606	59937.3
South Carolina	93.693	221045	252155
South Dakota	90.147	49557.9	55542.1
Tennessee	90.854	352461.2	376694.6
Texas	98.502	1815063.6	1556765.2
Utah	94.592	186910	171487.3
Vermont	98.66	30546.8	35126.6
Virginia	102.278	505351	486026
Washington	108.885	575129	454854.4
West Virginia	90.763	71343.2	82629.1
Wisconsin	93.347	306467.4	326703.2
Wyoming	91.418	36400	38263.6

Figure: Dataset used for the project

The dataset is obtained from the U.S. Bureau of Economic Analysis (BEA) and reflects prices across states as of December 2021.

The dataset of Regional Price Parity (RPP) contains 51 rows and 4 columns: States, RPP (Regional Price Parity), GDP (Gross Domestic Product) and Personal Income.

The RPP (Regional Price Parity) is a financial metric used to compare the cost of living in different American states and metro areas, GDP (Gross Domestic Product) measures the total value of goods and services produced within a country's borders and the personal income measures the total income received by individuals from all sources, including wages, salaries, and investment income.

1. **States** – The 50 States of the US and the District of Columbia: 51
2. **RPP** – Regional Parity Price of the state: 86.601 to 119.227
3. **GDP** – Gross Domestic Product of the state: 30546.8 to 2874730.8
4. **Personal Income** – Average Personal Income of the state: 35126.6 to 2332657.6

4. VISUALIZATIONS

4.1. BAR GRAPH

Used to visualize the Regional Price Parity by each state:

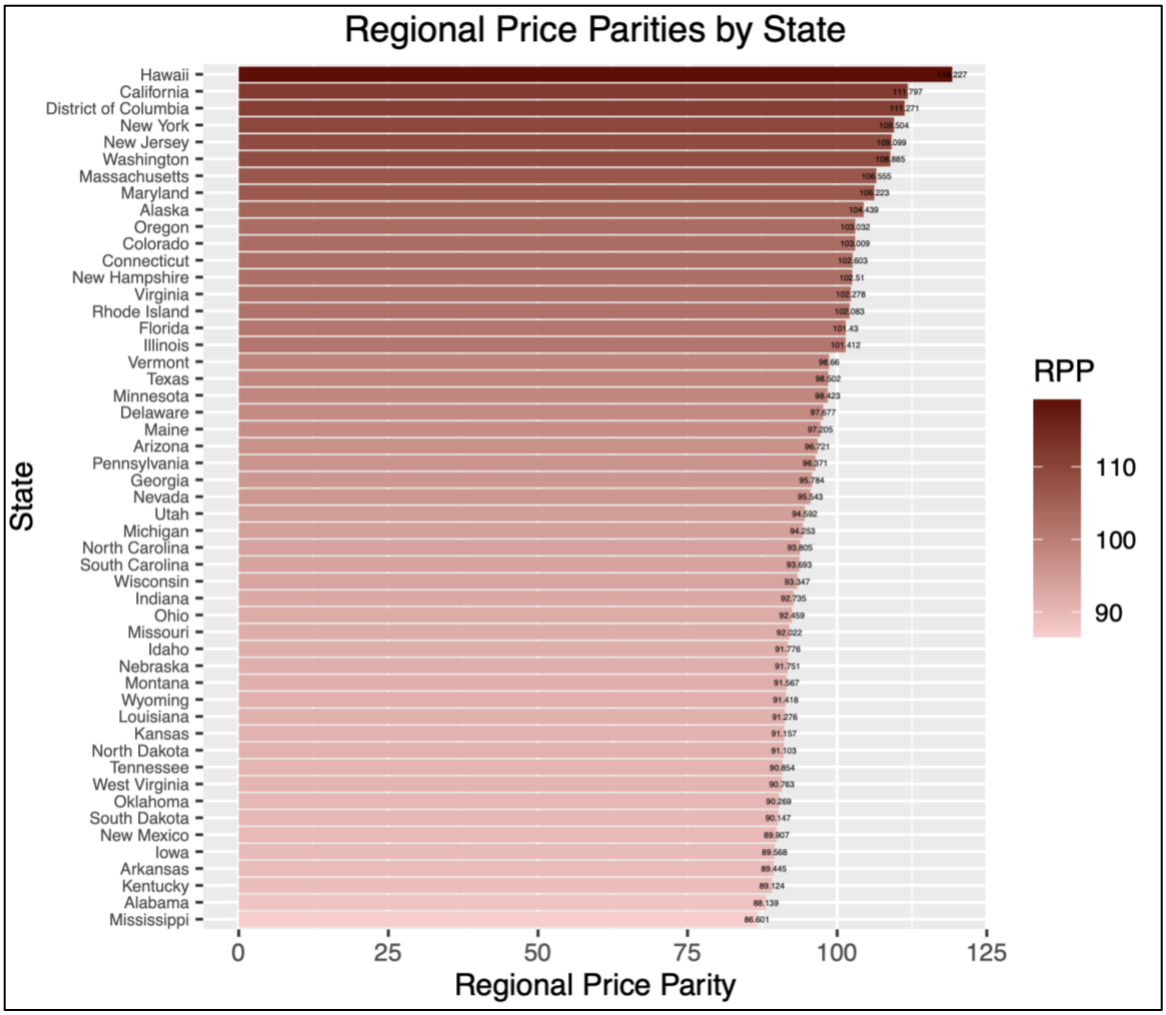

Figure: Bar Graph to visualize RPP by each state

The given graph depicts a dataset that ranks US states by their cost of living. The graph provides a clear comparison of the cost of living across the states and enables viewers to easily identify which states are more expensive or less expensive than the national average. Hawaii is ranked as the most expensive state with a price parity score of 119.3, which is 19.3% above the national average. Mississippi is ranked as the least expensive state with a price parity score of 84.4%, which is significantly below the national average.

4.2. MICROMAP

Used to plot the relationship between Gross Domestic Product (GDP) and the Average Personal Income:

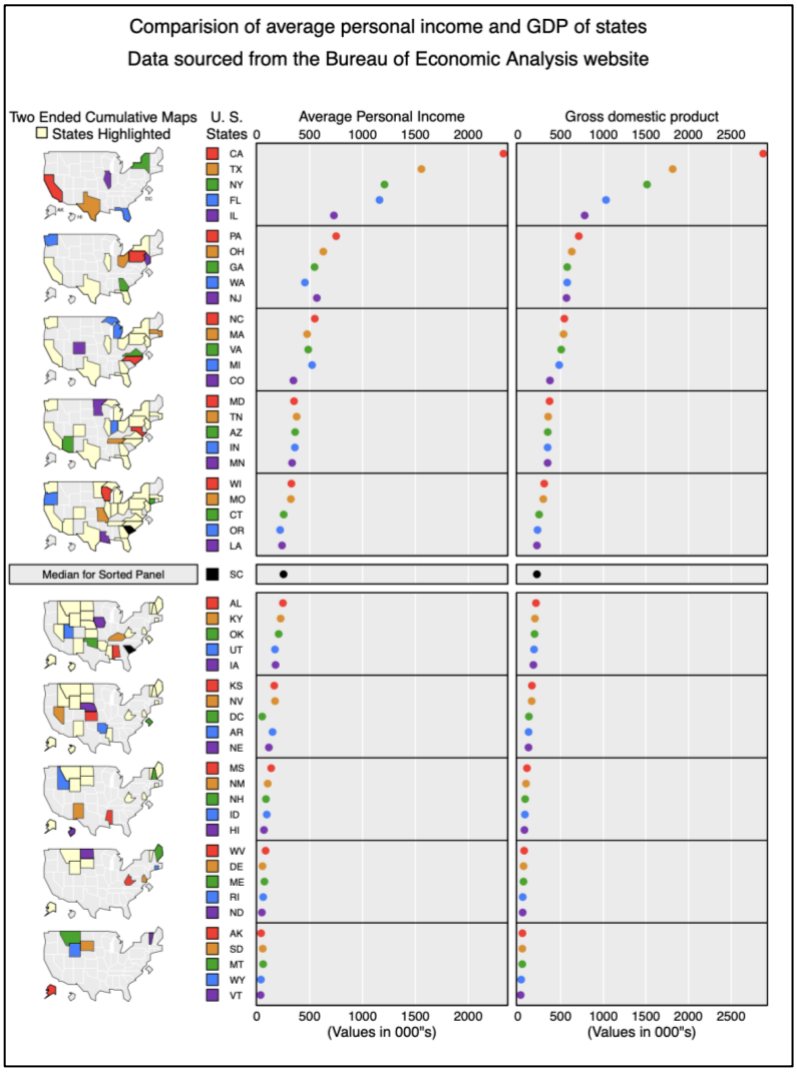

Figure: Micromap to visualize relationship between GDP and Average Personal Income

The graph shows the micromap of the GDP and the average personal income. GDP (Gross Domestic Product) measures the total value of goods and services produced within a country's borders. The personal income measures the total income received by individuals from all sources, including wages, salaries, and investment income. We can see that periods of strong GDP growth coincide with increases in personal income. Overall, plotting GDP and personal income data can provide valuable insights into the state of an economy and how it is evolving over time.

4.3. CORRELATION MATRIX AND HEATMAP

Used to plot the relationship between RPP AND GDP:

	RPP	GDP	Personal.Income
RPP	1.0000000	0.3941546	0.3560156
GDP	0.3941546	1.0000000	0.9912456
Personal.Income	0.3560156	0.9912456	1.0000000

Figure: Correlation Matrix

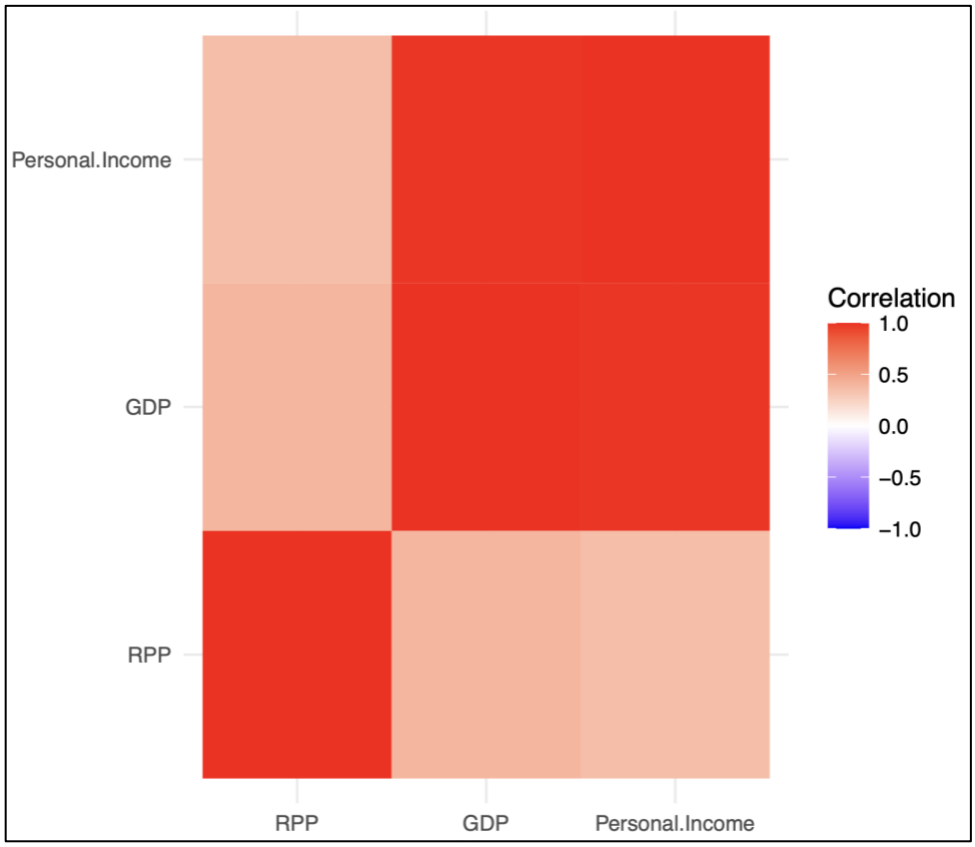

Figure: Heatmap to visualize RPP by each state

A correlation coefficient of 0.39 suggests a moderate positive correlation between RPP and GDP. This means that there is a tendency for higher values of RPP to be associated with higher values of GDP, but the correlation is not particularly strong. However, by using linear regression, we can explore the relationship between the two variables and gain additional insights into their relationship.

4.4. LINEAR REGRESSION ANALYSIS

Used to gain additional insights into relationship between RPP and GDP:

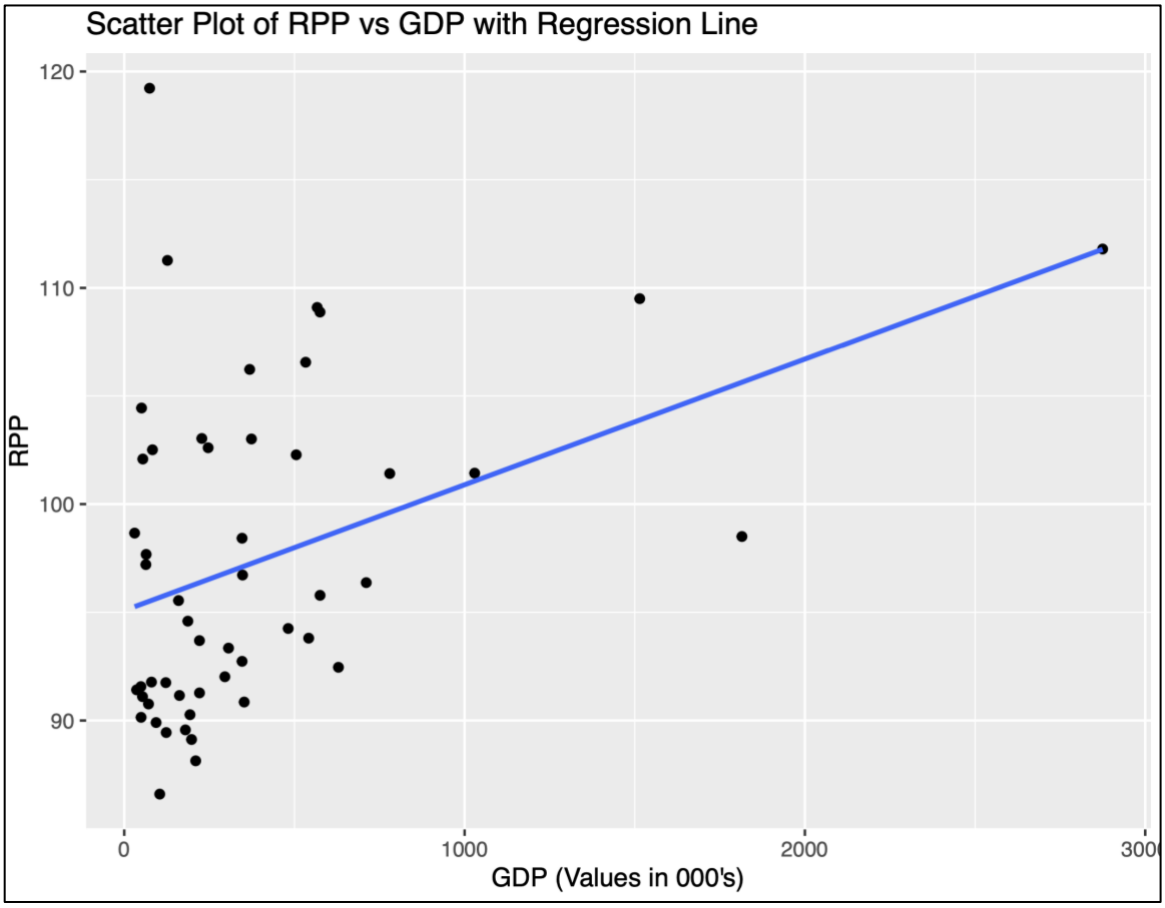

Figure: Scatter Plot of RPP vs GDP with Regression Line

This is a scatter plot of RPP (Regional Price Parity) vs GDP (Gross Domestic Product) with a regression line. The GDP values are represented by x-axis and the y-axis represents the RPP values. The regression line shows the relationship between the RPP and GDP values for the states.

Each point on the plot represents the RPP and GDP values for a particular state. Linear regression is used to model the relationship between GDP and RPP. The regression line is the straight line that best fits the data points and represents the average relationship between the two variables.
We can see that there is a positive correlation between the two variables. As GDP increases, so does RPP.

5. CONCLUSION

In conclusion, our analysis of the Regional Price Parity (RPP) dataset has provided valuable insights into the relationship between GDP and RPP, and how it varies across different regions and states within the United States.

By incorporating visualizations such as scatter plots, regression lines, and micromaps, we were able to uncover significant findings. The regression line, which represents the average relationship between GDP and RPP, revealed a positive slope. This indicates that as GDP increases, so does RPP, highlighting a correlation between economic output and the cost of living.

These findings hold substantial implications for various stakeholders, including businesses, policymakers, and individuals seeking a comprehensive understanding of economic conditions in different regions. By recognizing the positive correlation between GDP and RPP, decision-makers can make informed choices regarding investments, policy interventions, and personal financial planning.

6. REFERENCES

1. We have selected the dataset from https://howmuch.net/articles/regional-price-parities-by-state

2. We have sourced the data from https://www.bea.gov/data/prices-inflation/regional-price-parities-state-and-metro-area

APPENDIX

R SCRIPT:

```
library(dplyr)
library(ggplot)
library(ggplot2)
require("micromapST")
library(micromapST)
library(reshape2)
install.packages("reshape2")
library(reshape2)
install.packages("corrplot")

# Bargraph to visualize the Regional Price Parity by each state:
data <- read.csv("/Users/aakashboenal/Downloads/RPP.csv")
ggplot(data, aes(x=reorder(States, RPP), y=RPP, fill=RPP)) +
  geom_bar(stat="identity") +
  scale_fill_gradient(low="#ffcccc", high="#660000") +
  ggtitle("Regional Price Parities by State") +
  xlab("State") +
  ylab("Regional Price Parity") +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme(axis.text.y = element_text(size = 6))+
  geom_text(aes(label=RPP), size=1.0)+coord_flip()

# Micromap to plot the relationship between Gross Domestic Product (GDP) and the
Average Personal Income:
state_data <- read.csv("/Users/aakashboenal/Downloads/RPP.csv")
state_data
state_data$Personal.Income <- state_data$Personal.Income / 1000
state_data$GDP <- state_data$GDP / 1000
type <-c('maptail','id','dot','dot')
lab1 <-c(NA,NA,'Average Personal Income','Gross domestic product')
#lab2 <-c(NA,NA,'with 95% confidence intervals','with 95% confidence intervals')
lab3 <-c(NA,NA,'(Values in 000"s)','(Values in 000"s)')
col1 <-c('States','States','Personal.Income','GDP')
col2 <-c(NA,NA,'Personal.Income','GDP')
col3 <-c(NA,NA,'Personal.Income','GDP')
panelDesc <-data.frame(type,lab1,lab3,col1,col2,col3)
```

```
panelDesc
fName = "state_data.pdf"
pdf(file=fName,width=7.5,height=10)
micromapST( state_data,
        panelDesc,rowNamesCol ='States',
        rowNames ='full',sortVar ='GDP',
        ascend =FALSE,
        title =c("Comparision of average personal income and GDP of states","Data sourced from the Bureau of Economic Analysis website"),
        ignoreNoMatches =TRUE)
dev.off()

# Correlation Matrix
cor_matrix <- cor(state_data[,2:4])
cor_matrix
corr_df <- reshape2::melt(cor_matrix)

# Create the heatmap (using ggplot2) to plot the relationship between RPP AND GDP:
ggplot(corr_df, aes(x=Var1, y=Var2, fill=value)) +
  geom_tile() +
  scale_fill_gradient2(low="blue", high="red", midpoint=0, limit=c(-1,1), space="Lab", name="Correlation") +
  theme_minimal() +
  theme(axis.text.x = element_text(vjust = 1, size = 9),
       axis.text.y = element_text(size = 9))

# Create scatter plot of RPP vs GDP
ggplot(data, aes(x = GDP, y = RPP)) +
  geom_point() +
  labs(x = "GDP", y = "RPP") +
  ggtitle("Scatter Plot of RPP vs GDP")

# Add regression line to the scatter plot
ggplot(data, aes(x = GDP, y = RPP)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(x = "GDP", y = "RPP") +
  ggtitle("Scatter Plot of RPP vs GDP with Regression Line")
```